Expression of the Gene for the \( \beta \) Subunit of Mouse Thyrotropin Results in Multiple mRNAs Differing in Their 5'-Untranslated Regions*

(Received for publication, March 31, 1987)

Ofra Wolf, Ione A. Kourides, and James A. Gurr

From the Laboratory of Molecular Endocrinology, Memorial Sloan-Kettering Cancer Center and Cornell University Medical College and Graduate School of Medical Sciences, New York, New York 10021

The gene encoding the \( \beta \) subunit of mouse thyrotropin (TSH\( \beta \)) has been isolated from a mouse genomic library, and its nucleotide sequence has been determined. Blot hybridization analysis of restriction enzyme digests of mouse DNA indicates that there is a single mouse TSH\( \beta \) gene. The gene is 4.8 kilobases in length and contains five exons, which are 27, 47, 41, 163, and 328 base pairs long. Exons 1, 2, and 3 encode only 5'-untranslated mRNA sequences and are separated by introns that are 150 and 380 base pairs long. The protein-coding mRNA sequences are found in exons 4 and 5 and are interrupted by a 460-base pair intron. The position of this intron, between the codons for amino acids 34 and 35, has been conserved in all the known glycoprotein hormone \( \beta \) subunit genes. Exons 3 and 4 are separated by a large 3.2-kilobase intron. When primer extension analysis, using an oligonucleotide primer complementary to exon 4 sequences, was employed to locate the transcription start site, four products were obtained. Nucleotide sequencing of these products showed that they were derived from separate TSH\( \beta \) mRNAs that differed in the lengths of their 5'-untranslated regions. These 5'-untranslated mRNA sequences are derived from different combinations of exons 1, 2, and 3, each spliced to exons 4 and 5. The longest 5'-untranslated mRNA sequence, 116 nucleotides long, includes exons 1, 2, and 3, and the first base of exon 4; the shorter 5'-untranslated regions, 75, 69, and 28 nucleotides long, arise by splicing out the second and/or the third exon sequences.

In contrast to the mouse TSH\( \beta \) gene, transcription of the rat TSH\( \beta \) gene from the analogous start site has been reported to give only a single mRNA, with a 5'-untranslated region of 28 nucleotides. Divergence of the mouse and rat TSH\( \beta \) gene sequences at RNA splice sites can account for the absence of exon 2, but not exon 3, sequences in rat TSH\( \beta \) mRNA. Primer extension and RNase protection analyses also showed that the mouse TSH\( \beta \) gene contains a second transcription start site, located 43 base pairs upstream of the first start site, in a position corresponding to that in the rat TSH\( \beta \) gene. Each start site in the mouse gene is flanked by characteristic TATAA box and CAAT box sequences. In the hypothroid mouse pituitary and in mouse thyrotropic tumors, transcription occurs predominantly from the downstream start site. Elucidation of the structure of the mouse TSH\( \beta \) gene and its 5'-flanking region facilitates the study of the mechanism controlling the production of multiple mouse TSH\( \beta \) mRNAs and the use of the alternative promoters.

TSH, the pituitary hormone that is the major regulator of thyroid hormone production by the thyroid gland (1), is one of the family of glycoprotein hormones, which includes pituitary follicitropin and LH and placental CG (2). Each of these hormones consists of two dissimilar, noncovalently associated subunits, \( \alpha \) and \( \beta \). Within a species, these hormones share a common \( \alpha \) subunit; the biological specificity of the complete hormones is determined by the unique \( \beta \) subunits (2). The \( \alpha \) subunit and the pituitary \( \beta \) subunits are each encoded by a single gene (3-12), each of which is located on a different chromosome (13-17). In contrast, there are six closely linked human CG\( \beta \) genes arranged in tandem and inverted pairs (18-21).

The production of TSH is suppressed by thyroid hormones. We and others (22-25) have shown previously that, in mouse thyrotropic tumors and in the mouse and rat hypothyroid pituitary, thyroid hormones rapidly inhibit \( \alpha \) and TSH\( \beta \) gene transcription and that inhibition of TSH\( \beta \) gene transcription is much more marked than that of \( \alpha \). The data are consistent with a mechanism involving the direct interaction of the nuclear thyroid hormone-receptor complex with regulatory regions of the \( \alpha \) and TSH\( \beta \) genes to modulate transcription.

In order to delineate the sequence elements that mediate thyroid hormone action on TSH\( \beta \) gene transcription, we have isolated and sequenced genomic fragments containing the structural and 5'-flanking regions of the mouse TSH\( \beta \) gene, using as a hybridization probe a mouse TSH\( \beta \) cDNA we have described previously (26). We have found that there is a single TSH\( \beta \) gene in the mouse and that this gene contains two sites for the initiation of transcription, as does the rat TSH\( \beta \) gene (11). Transcription of the mouse TSH\( \beta \) gene from the predominant start site gives rise to multiple TSH\( \beta \) mRNAs that differ in the lengths of their 5'-untranslated regions. The 5'-untranslated region of the longest of the major mouse TSH\( \beta \) mRNAs is encoded by three exons; shorter mRNAs are formed by splicing out the second and/or the third exons. In contrast,

---

*This work was supported by United States Public Health Service Grants CA 23185 and CA 08748 and by Grant BC-502 from the American Cancer Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03450.

† Present address: Dept. of Plant Genetics and Breeding, Inst. of Seed and Garden Crops, Agricultural Research Organization, Volcani Center, Bet Dagan, 50250, Israel.

§ To whom correspondence should be addressed: Head, Laboratory of Molecular Endocrinology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021.

1 The abbreviations used are: TSH, thyrotropin; LH, lutropin; CG, chorionic gonadotropin; bp, base pair(s); kb, kilobase(s); PIPES, 1,4-piperazinediethanesulfonic acid; UTP, uridine 5'-triphosphate.
only a single TSHβ mRNA has been reported for transcription of the rat TSHβ gene from the corresponding start site.

**EXPERIMENTAL PROCEDURES**

**Materials—**Restriction endonucleases and DNA-modifying enzymes were purchased from Boehringer Mannheim, Bethesda Research Laboratories, and New England Biolabs and were used as recommended by the suppliers. Radioactive nucleotides, [γ-32P]ATP (3000 Ci/mmol), [α-32P]dATP, [α-32P]dCTP, [α-32P]dGTP, [α-32P]UTP, and [α-32P]TTP (800 Ci/mmol each), and [α-32P]cordycepin (5000 Ci/mmol) were purchased from Du Pont-New England Nuclear. Nitrocellulose and DEAE-nitrocellulose paper (NA45) were from Schleicher & Schuell.

**Screening Mouse Genomic Libraries—**Two BALB/c mouse genomic libraries, a partial HaeIII library and a partial EcoRI digest, cloned in the bacteriophage λ Charon 4A, were kindly provided by Dr. Leroy Hood (California Institute of Technology). They were screened by the method of Benton and Davis (27) using as a hybridization probe a mouse TSHβ cDNA plasmid (26) labeled with [α-32P]dCTP by nick translation (28). DNA was isolated from positive clones after repeated plaque purification (29) and was characterized by restriction enzyme mapping and Southern blotting (30).

**Subcloning of Phage DNA Restriction Fragments—**To facilitate DNA sequencing, a 3.4-kb BamHI-EcoRI and a 1.8-kb EcoRI-HindIII fragment were subcloned into the plasmid pBR322. Plasmid DNA was purified by CsCl gradient centrifugation (29).

**Isolation and Analysis of Genomic DNA—**High molecular weight genomic DNA was prepared from livers of BALB/c, LAF, and CD, mice by the method of Gross-Bellard et al. (31) and digested with restriction enzymes under standard conditions. After electrophoresis on 1% agarose gels, the DNA was blotted onto nitrocellulose filters and transferred to nitrocellulose as described by Thomas (35).

**DNA Restriction Fragments—**To facilitate DNA sequencing, a 3.4-kb BamHI-EcoRI and a 1.8-kb EcoRI-HindIII fragment were subcloned into the plasmid pBR322. Plasmid DNA was purified by CsCl gradient centrifugation (29).

**DNA Sequencing—**DNA sequencing was performed by the chemical degradation method of Maxam and Gilbert (36). Restriction fragments were labeled at the 5′-end with [γ-32P]ATP and T4 polynucleotide kinase (29) and at the 3′-end with either the Klenow fragment of Escherichia coli DNA polymerase I and the appropriate [α-32P]dNTP (29) or with [α-32P]cordycepin and terminal deoxynucleotidyltransferase (37). Labeled fragments were separated by polyacrylamide gel electrophoresis and isolated using an electroeluter (International Biotechnologies, Inc.). Each sequence was determined in at least two independent experiments.

**RNA Isolation—**RNA was isolated from the pituitaries of LAF, hypothyroid mice, mouse TSH-secreting tumors (38), and mouse liver by homogenization in guanidinium thiocyanate and centrifugation through CsCl (39). The RNA pellet was redissolved in 10 mM Tris-HCl, pH 7.4, 5 mM EDTA, 1% sodium dodecyl sulfate, extracted with phenol/chloroform, precipitated with ethanol, and redissolved in water.

**Primer Extension Analysis—**Oligonucleotide-primed cDNA syn-
thesis with reverse transcriptase was used to determine the position of the start site of transcription of the mouse TSHp gene and to characterize the 5'-untranslated region of the TSHp mRNAs. Oligonucleotides were obtained from the Beckman Research Institute, City of Hope Medical Center, Duarte, CA. They were synthesized by the solid phase phosphotriester method on a Systec Microsyn 1450 automated DNA synthesizer and purified using a PRP-1 reverse phase column and a linear gradient of 5-35% acetonitrile (40). Oligonucleotides were labeled at the 5'-end with \([\text{Y}^32\text{P}]\)ATP and polynucleotide kinase to a specific activity of 2-10 \(\times 10^6\) cpm/pmol (29).

0.3 pmol of labeled primer was annealed to 25-50 pg of total RNA from a mouse thyrotropic tumor, mouse liver, or yeast tRNA and to 10-25 pg of mouse hypothyroid pituitary total RNA by incubation in a 10-pl reaction containing 20 mM Tris-Cl, pH 8.3, 50 mM KCl, 0.1 mM EDTA, 25 units of RNasin (Promega) at 70 °C for 5 min and allowing the mixture to cool to 42 °C. The reaction mixture was then adjusted, in a final volume of 20 pl, 50 mM Tris-Cl, pH 8.3, 50 mM KCl, 10 mM dithiothreitol, 10 mM MgCl₂, and 1 mM each of dATP, dCTP, dGTP, and dTTP. Reverse transcription was initiated by addition of 25 units of avian myeloblastosis virus reverse transcriptase. Incubation was for 1 h at 42 °C, and the reaction was terminated by addition of EDTA to 10 mM. After phenol/chloroform extraction, the reaction products were precipitated with ethanol, redissolved in Loading Buffer (80% formamide, 50 mM Tris borate, pH 8.3, 1 mM EDTA, 0.1% xylene cyanol, 0.1% bromphenol blue), and separated by electrophoresis on polyacrylamide/urea sequencing gels (31). Labeled products were visualized by exposure of the gels to Kodak XAR-5 x-ray film at -70 °C with an intensifying screen. The size of the labeled bands was estimated by co-electrophoresis with \(32\text{P}\)-end-labeled restriction enzymes or oligonucleotides. For DNA sequence analysis, primer extension products were electroeluted from the gels and sequenced by the chemical degradation method (36).

**RESULTS**

**Isolation of Genomic Fragments Containing the Mouse TSHβ Gene**—A mouse genomic library consisting of a partial Atul + HaelIII digest of BALB/c mouse DNA cloned in λ Charon 4A was screened using a \(32\text{P}\)-labeled mouse TSHβ plasmid probe, which we have described previously (26). This probe contains the entire TSHβ mRNA protein-coding region, all of the 3'-untranslated region, and 56 bp of 5'-untranslated mRNA sequence. Screening of 1.2 \(\times 10^8\) plaques yielded 14...
FIG. 4. Determination of the transcription start site of the mouse TSHβ gene by primer extension. An oligonucleotide primer complementary to nucleotides 2–18 in exon 4 (underlined in Fig. 3) was labeled at its 5'-end with 32P and hybridized to total RNA from a mouse thyrotropic tumor (lane A), mouse hypothyroid pituitary (lane B), or mouse liver (lane C). The primer was then extended with reverse transcriptase, and the reaction products were separated by electrophoresis on a 10% polyacrylamide/urea sequencing gel. The autoradiogram was obtained after 19 h exposure to x-ray film. Lane M, 3'-end-labeled fragments of an MspI digest of pBR322 were used as size markers. Sizes are in base pairs. The arrows indicate cDNAs I–IV, which were excised from the gel for nucleotide sequencing (Fig. 5).

positives. After plaque purification, restriction enzyme analysis showed that 12 of the phage inserts were identical, represented by λ24, whereas each of the other two, λ9 and λ90, was unique (Fig. 1, A and B). The analysis also showed that the genomic DNA fragments in λ9 and λ90 overlapped, and they were therefore assumed to be derived from the same TSHβ gene. Although λ24 shared the same structure as λ9 and λ90 in its 3'-region, its 5'-region was unique. We initially interpreted this as suggesting the existence of two different mouse TSHβ genes (43). However, further work indicated that this was not the case. For example, screening of a second BALB/c genomic library yielded six positive plaques, none of which contained DNA inserts characteristic of λ24. Southern blot analysis of genomic BALB/c DNA showed only hybridizing bands predicted from the structures of λ9 and λ90 DNA and none characteristic of λ24. Also, hybridization with a probe specific for the 5'-noncoding region of the TSHβ cDNA revealed that only λ9 and λ90, and not λ24, contained TSHβ 5'-exon sequences. We therefore concluded that the 5'-portion of the genomic DNA insert of λ24 (the stippled region in Fig. 1A) was not a normal part of the mouse TSHβ gene region and that its occurrence in λ24 adjacent to TSHβ gene sequences was due to DNA rearrangement during construction or amplification of the library.

Analysis of Mouse Genomic DNA.—To determine the number of TSHβ genes in the mouse, high molecular weight DNA was prepared from BALB/c mouse liver and analyzed by restriction enzyme digestion, followed by hybridization of Southern blots with 32P-labeled mouse TSHβ cDNA plasmid probe. The results for three restriction enzyme digests are shown in Fig. 2. The probe hybridized with fragments of approximately 8.3 and 1.8 kb in the EcoRI + HindIII digest, 4.6, 2.2, and 0.5 kb in the SalI digest, and 3.6 and 1.4 kb in the PstI digest. Analysis of genomic DNA from LAF, and CD1 mice showed similar simple patterns of restriction enzyme fragments (Fig. 2), suggesting that there was a single TSHβ gene in the mouse. The pattern of hybridizing fragments obtained in the analysis of the BALB/c genomic DNA was consistent with the restriction map of the cloned TSHβ gene (Fig. 1C).

Sequencing of the Mouse TSHβ Gene.—The complete TSHβ cDNA plasmid probe hybridized to a 1.8-kb EcoRI-HindIII fragment in λ9 and λ90 (Fig. 1B). This fragment was subcloned into pBR322, mapped in detail, and its nucleotide sequence determined (Fig. 1C and Fig. 3). The fragment contained two exons (Fig. 1C, E4 and E5), 163 and 328 bp long, separated by a 460-bp intron (Fig. 1C, IVS D). These exons encode the entire TSHβ pre-protein, the 3'-untranslated mRNA sequences, and one nucleotide of 5'-untranslated mRNA. The intron interrupts the protein-coding mRNA sequence between the codons for amino acids 34 and 35. To locate the exon(s) encoding the remaining 5'-untranslated mRNA, a 117-bp AluI fragment containing 73 bp of this region was excised from the TSHβ cDNA and used as a probe. This probe did not hybridize with the subcloned EcoRI-HindIII fragment but did hybridize to λ9 and λ90 about 3 kb upstream of the protein-coding exons. A hybridizing 3.4-kb BamIII-EcoRI fragment was then subcloned into pBR322, mapped, and sequenced (Fig. 1, B and C, and Fig. 3). It contained the remaining known 88 bp of 5'-untranslated mRNA sequence in two exons (Fig. 1C, E2 and E3), 47 and 41 bp long, respectively, separated by a 380-bp intron (Fig. 1C, IVS B). These two exons are separated from the protein-coding exons by a large intron of approximately 3.2 kb (Fig. 1C, IVS C). The nucleotide sequence of the TSHβ gene subclones and the 5'- and 3'-flanking regions is shown in Fig. 3.

Determination of the Site of Initiation of Transcription.—The position of the transcription start site of the mouse TSHβ gene was determined by primer extension analysis. An oligonucleotide primer complementary to the first 18 bases of the mRNA protein-coding sequence (Fig. 3) was labeled at its 5'-end, annealed to total RNA isolated from a mouse TSH-secreting tumor, and extended by incubation with reverse transcriptase. When the reaction products were analyzed by denaturing polyacrylamide gel electrophoresis, four major labeled cDNAs were seen (Fig. 4, lane A), with lengths of 132–135 (band I), 93 (band II), 84 (band III), and 44–47 (band IV) nucleotides. Identical bands were obtained with hypothyroid mouse pituitary total RNA (Fig. 4, lane B), showing that the multiple primer extension products were not due to any abnormal properties of the tumor. The specificity of the primer for TSHβ mRNA was shown by the fact that none of these reaction products was obtained using total RNA from mouse liver (Fig. 4, lane C), a tissue that does not express TSHβ.

The multiple primer extension products could have resulted from the premature termination of cDNA synthesis by reverse transcriptase along a single TSHβ mRNA. Alternatively, they
could have been full length transcripts arising from hybridization of the primer to multiple TSHβ mRNAs that differed in the lengths of their 5'-untranslated regions. The size of the longest primer extension product (band I), 132-135 nucleotides, was consistent with a TSHβ mRNA with a 5'-untranslated region of about 115 nucleotides. This implied that the sequence of the TSHβ cDNA which we had used to delineate the exon sequences in TSHβ gene was incomplete and that the position of about 25 bp of exon sequences remained to be determined. To distinguish between the two possible origins of the multiple cDNA bands and to obtain the complete sequence of the TSHβ mRNA 5'-untranslated region(s), cDNAs I-IV were eluted from a denaturing gel, and their nucleotide sequences were determined by the chemical degradation method. The autoradiograms shown in Fig. 5 give the nucleotide sequences of cDNAs I-IV immediately upstream of the oligonucleotide primer and thus show the exon sequences which are spliced to exon 4. The sequences of cDNAs II and IV are different in this region, and each is different from the sequence shared by cDNAs I and III. The complete sequences of the latter cDNAs showed that they in turn diverged further upstream. However, at their 5'-end each of the four cDNAs shared the same sequence, that of cDNA IV. These data demonstrated that each cDNA was derived by hybridization of the primer to an mRNA with a different 5'-untranslated region. None of the sequences contained in cDNA IV was found in the TSHβ cDNA which we had previously characterized, and comparison with the sequence of the TSHβ gene showed that the cDNA IV sequence was derived from a separate exon (Fig. 1, ET) which we had not identified previously. We inferred that, because this exon was found at the 5' terminus of each of the different mRNAs, it also contained the transcription start site. Thus, the complete mouse TSHβ gene contains five exons, three of which encode 5'-untranslated mRNA sequences (Fig. 3).

The sequence data showed that the mRNAs corresponding to cDNA bands I-IV contained 5'-untranslated regions derived from different combinations of exons 1, 2, and 3, each spliced to exons 4 and 5. Each 5'-untranslated region also contained the single untranslated base at the 5'-end of exon 4. In addition to this single nucleotide, the 5'-untranslated mRNA region corresponding to cDNA I (114-117 nucleotides long) is coded for by exons 1, 2, and 3, cDNA II (75 nucleotides) by exons 1 and 2, cDNA III (69 nucleotides) by exons 1 and 3, and cDNA IV (26-29 nucleotides) by exon 1 alone. From the size of the primer extension products, exon 1 was estimated to be 26-29 bp long, the ambiguity arising from the fact that a doublet was obtained with each cDNA. This heterogeneity may reflect either a true heterogeneity in the mRNA population or may be the result of interference of the 3' cap site in the action of reverse transcriptase. Densitometry of the autoradiogram in Fig. 4 showed that mRNAs corresponding to bands I-IV constituted approximately 55, 20, 20, and 5% of the total TSHβ mRNA, respectively, in the tumor and in the hypothyroid pituitary.

To determine the position of the start site more exactly, a second primer extension experiment was performed (Fig. 6) using a 19-mer oligonucleotide primer complementary to bases 9-28 in exon 1 (Fig. 3). Hybridization of labeled primer with total RNA from both a mouse thyrotropic tumor (Fig. 6, lane A) and hypothyroid pituitary (lane B), followed by extension with reverse transcriptase, gave major products 25

**Fig. 5. Nucleotide sequencing of primer extension products.** The nucleotide sequences of cDNAs I-IV (Fig. 4) were determined by the chemical degradation method. The autoradiogram shows the sequence ladders obtained after electrophoresis on a 20% polyacrylamide/urea sequencing gel. The nucleotide sequence shown on the right of each panel is that of the mRNA coding strand of the gene, the complement of the sequence shown in Fig. 3. The arrows denote the 5'-end of the oligonucleotide primer.
and 27 nucleotides long. Specificity of the primer for TSHβ mRNA was shown by the absence of these bands after hybridization with mouse liver RNA (lane C) and tRNA (lane D). The nucleotide sequence of the two cDNAs was shown to be identical, with the addition of two bases in the longer form. The synthesis of a doublet of cDNAs was again probably due to the cap structure.

Allowing for this ambiguity, from the combined results of the primer extension analyses with the two different probes, the start site of transcription of the mouse TSHβ gene was assigned to the adenine residue 116 bp upstream of the start site of translation (Fig. 3), giving exon 1 a length of 27 bp.

Longer exposure of the gels obtained using the primer from exon 4 revealed an additional band at 182 nucleotides (data not shown). The size of this band suggested the existence of a second transcription start site about 40 bp upstream of the major start site used to synthesize mRNAs I-IV. Consistent with this was the finding of a minor band of 69 bp with the exon 1 primer (data not shown). Further evidence for a second transcription start site for the mouse TSHβ gene was obtained from the RNase protection studies described below.

**RNase Protection Analysis**—The position of the transcription start site was also determined by RNase protection analysis (Fig. 7). A uniformly labeled 261-bp TaqI-PstI RNA probe containing exon 1 and about 80 bp of 5'-flanking sequence (Fig. 3) was synthesized using SP6 RNA polymerase and [α-32P]UTP (Fig. 7, lane A). Hybridization of this probe with tumor (Fig. 7, lane B) and hypothyroid pituitary (lane C) total RNA followed by digestion of the products with RNase A and RNase T1 gave a major band 25 nucleotides long, in both cases. Specificity was shown by the absence of this band after hybridization with total mouse liver RNA (lane D) and tRNA (lane E). The size of this band confirmed the assignment of the position of the cap site based on the primer extension data. There was also a band at 70 nucleotides with both tumor and pituitary RNA (lanes B and C), which we attributed to the TSHβ mRNA with a longer form of exon 1, which was suggested by the primer extension analyses. The size of this product implied that the mouse TSHβ gene contained a second transcription start site that was located 43 bp upstream of the major start site (Fig. 3).

The autoradiogram also showed a protected band about 180 nucleotides long. The size of this band is consistent with results obtained in an earlier report of an S1 nuclease analysis in which we described a TSHβ mRNA with a 5'-untranslated region of 226 nucleotides (44). The data suggest that this minor TSHβ RNA contains a 5'-untranslated region consisting of exon 1, intron A, exon 2, and exon 3. It may be either a TSHβ mRNA in which intron A has not been spliced out or a TSHβ nuclear precursor RNA.

**DISCUSSION**

Using a mouse TSHβ cDNA probe, we have isolated recombinant phage containing overlapping DNA fragments encoding the entire mouse TSHβ subunit gene. Southern blotting of restriction enzyme digests of mouse genomic DNA indicated that there is a single TSHβ gene in the mouse, as in the human (9, 12) and the rat (10, 11). The mouse TSHβ gene is 4.8 kb long and contains five exons (Fig. 1C). Exons 4 and 5 encode the mRNA protein-coding region, including the signal peptide and a single nucleotide of 5'-untranslated mRNA. Exons 1, 2, and 3 encode only 5'-untranslated mRNA sequences. The major portion of the gene consists of the long 3.2-kb intron (IVS 3) which separates the two groups of exons. Although the mouse TSHβ gene is similar in size and organization to the rat gene (10, 11), the latter differs in

**Fig. 6. Determination of the size of exon 1 of the mouse TSHβ gene by primer extension.** An oligonucleotide primer complementary to bases 9-28 in exon 1 (underlined in Fig. 3) was 5'-end-labeled and hybridized to total RNA from mouse thyrotropic tumor (lane A), mouse hypothyroid pituitary (lane B), mouse liver (lane C), and tRNA (lane D). The primer was then extended with reverse transcriptase, and the reaction products were separated by electrophoresis on a 12% polyacrylamide/urea sequencing gel. The autoradiogram was exposed to x-ray film for 4 h. Lane M, 5'-end-labeled oligonucleotide size markers (nucleotides).

**Fig. 7. Determination of the size of exon 1 of the mouse TSHβ gene by RNase protection.** The PstI-TaqI gene fragment spanning exon 1 (Fig. 2) was uniformly labeled with [α-32P]UTP and SP6 RNA polymerase after insertion into the plasmid SP64. The probe was hybridized with total RNA from a mouse thyrotropic tumor (lane B), mouse hypothyroid pituitary (lane C), mouse liver (lane D), and tRNA (lane E), the reaction mixtures were digested with RNase A and RNase T1, and the products separated by electrophoresis on a 10% polyacrylamide/urea sequencing gel. The autoradiogram was obtained after 64 h exposure to x-ray film. Lane A, undigested probe; lane M, size markers (base pairs) from a 3'-end-labeledMspI digest of pBR322; lane M, 5'-end-labeled oligonucleotide size markers (nucleotides).
having its 5'-untranslated region encoded by only a single exon, which corresponds to exon 1 in the mouse gene. The genes for the rat (6), bovine (7), and human (18-21) LβH and the human CGβ (18-21) subunits have a short intron between the codons for amino acids 41 and 42. The position of this intron is identical to that in the mouse, rat, and human TSHβ genes with respect to the subunit amino acid sequences because, in all of these genes, it falls nine nucleotides downstream of the codons for the completely conserved element Cys-Ala-Gly-Tyr-Cys. Each of the known LβH and CGβ genes (6, 7, 18-21) has a second short intron between the codons for amino acids -15 and -16 in the signal sequence. The corresponding intron in the mouse and rat TSHβ genes is similarly placed, 15 nucleotides upstream, one nucleotide into the 5'-untranslated mRNA region. The length of this intron in the TSHβ genes accounts for the large difference in size between the TSHβ subunit genes (4.8 kb long) and the LβH and CGβ genes (about 1 kb long). However, the overall structure of the glycoprotein hormone β subunit genes has been fairly well conserved.

The mouse TSHβ gene possesses two transcription start sites (Fig. 3). Examination of the gene sequences flanking each of the start sites shows the sequence TATAAA beginning 29 and 24 bp upstream of positions +1 and +1', respectively (Fig. 8). This "TATAA box" consensus sequence is found 25–30 bp upstream of many other eukaryotic transcription start sites and is thought to determine the correct position of initiation of transcription by RNA polymerase II (45). In addition, consensus "CAAT box" sequences (45) are located beginning at positions -73 and -74 with respect to the cap site at +1 and at -73 with respect to the site at +1'. It has been proposed that this sequence element increases the efficiency of transcription, and it is found typically 70–90 bp upstream of the cap site in many eukaryotic genes (45).

The striking feature of the mouse TSHβ gene is that expression from the downstream start site results in multiple mRNAs differing in the lengths of their 5'-untranslated regions. In contrast, although the rat TSHβ gene has a transcription start site in the same position as that in the mouse gene (Fig. 8A), transcription from this start site in the rat gives rise to only a single mRNA (11). This rat TSHβ mRNA has a 5'-untranslated region 28 nucleotides long and therefore corresponds to band IV seen in the mouse (Fig. 4). In the hypothyroid mouse pituitary, this mRNA constitutes only 5% of the total TSHβ mRNA (Fig. 4). In addition, the TSHβ mRNAs with 5'-untranslated regions corresponding to cDNAs I, II, and III in the mouse are not expressed in the rat. The comparison of the sequences of exons 1, 2, and 3 and their flanking regions in the mouse and the rat TSHβ genes in Fig. 8 shows the strong sequence homology between the two genes. However, the lack of the equivalent of mouse exon 2 sequences in rat TSHp mRNA can be attributed to sequence divergence. The dinucleotides GT and AG are completely conserved at 5' and 3' intron termini, respectively, of virtually all vertebrate genes (45–47). However, at the site in the rat TSHβ gene corresponding to the splice junction between intron A and exon 2 in the mouse gene, there is an AA pair rather than an AG pair (Fig. 8B). Similarly, at the position in the rat TSHβ gene analogous to the exon 2-intron B splice junction in the mouse gene there is a GC rather than a consensus GT dinucleotide (Fig. 8B). Thus, the 47 bases in the rat gene which are analogous to those expressed as exon 2 in the mouse gene are not flanked by essential RNA splicing signals and cannot be processed into mature mRNA. A similar argument cannot account, however, for the lack of expression of exon 3 sequences in rat TSHβ mRNA. Potential splicing signals are present in the rat RNA precursor (Fig. 8C), but they are apparently not used.

The functional significance of the multiple mRNAs generated by alternative splicing of the transcript from the major start site in the mouse TSHβ gene is unclear, because an identical protein is produced in each case. More commonly, variable splicing of a single RNA transcript involves the inclusion or exclusion of exon sequences in mature mRNAs so as to alter protein structure (reviewed in Refs. 47 and 48). The properties of the related proteins may be significantly different such that expression of the alternative mRNAs is tissue-specific or developmentally regulated (49–51).
As in the mouse TSHβ gene, the rat gene has an alternative transcription start site (11), and transcription from this second site gives a rat TSH mRNA with a 5′-untranslated region of 71–78 nucleotides (10, 11). Use of these alternative promoters again generates different mRNAs but identical proteins, and the factors governing promoter choice are unknown. For other genes, tissue specificity of promoter choice has been shown (52–55). For example, the mouse α-amylase gene uses separate liver- and salivary gland-specific promoters to produce an identical protein in each tissue (53). In such cases, the different lengths of the 5′-untranslated regions of multiple mRNAs may alter some property of the mRNAs which accounts for the pattern of their expression. For example, there may be significant differences in mRNA stability or translatability between the various forms. Differential expression from alternative start sites may be determined by relative promoter strength, by the presence of cis regulatory elements, or by trans-acting factors. Carr et al. (11) have recently reported that only expression from the downstream start site of the rat TSHβ gene is modulated by thyroid hormones. In contrast, we have observed that thyroid hormone regulates transcription from both start sites in the rat TSHγ gene. Presumably, there are promoter-specific regulatory sequences flanking the two start sites in the rat TSH gene which determine their sensitivity to thyroid hormones.

The isolation of the mouse TSHβ gene and the determination of its structure now allow us to elucidate the sequence elements involved in the regulation of TSHβ gene expression. Our finding of alternative transcription start sites and the generation of multiple TSHβ mRNAs from a single gene adds multiple TSHβ mRNAs from a single gene adds multiple TSHβ mRNAs from a single gene adds

Acknowledgments—We thank Andrew Goldschmidt, Cynthia Wagner, and Jonathan Norcross for excellent technical assistance and Drs. G. Kerr Whitfield, Peter T. Gudjon, James F. Catterall, and Glen L. Ginsalus for helpful discussions.

REFERENCES
